**DATA MANAGEMENT AND SHARING PLAN**

An example from an application focusing on secondary data analysis on data from human subjects.

**Data Type**

The data to be shared will include MRI images and clinical assessments from human research participants. This application is focused on secondary data analysis from existing data but will also deposit privately held data to a public repository. The existing data is available from the NIMH Data Archive (NDA) in collections 2134 (148 subjects) and 2433 (47 subjects).

In addition, we have data from a previous study involving 155 research participants with major depressive disorder that have not yet been shared with the research community but will be uploaded to NDA during the second quarter of the first year of funding.

As discussed in the application, structural MRI scans are available for time points before and after treatment along with relevant clinical data. Preparation for submitting that data to NDA is largely complete. Typical metadata necessary to link records within the dataset will be included, although the policies of our institution mandate that exact dates will not be shared (see Access section).

**Related Tools, Software, and/or Code**

The basic statistical analyses described in the application will be done using R. We plan to use the MRI data analysis tools in the FMRIB Software Library (FSL) for multi-level modeling of group effects. BrainVoyager software will be used for anatomical segmentation to isolate regions of interest within individual subjects, and the AI-powered analyses described in the application will use custom code written with the PyTorch library for Python. R, FSL, Python, and PyTorch are all freely available to the research community. [BrainVoyager](https://www.brainvoyager.com/) is commercial software, with licenses available for purchase.

All R and Python code (including trained model weights) will be available on our [lab bitbucket page](https://www.thelab.com/) no later than when publications are submitted.

**Standards**

The data that will be used for some of the proposed secondary data analysis is already in NDA and is formatted using NDA data dictionaries. The new data we will deposit will also use existing NDA data dictionaries. Since the data set to be deposited into NDA was collected prior to the publication of NOT-MH-20-067, not all of the common data elements expected by NIMH are available. However, we will transform some existing demographic and clinical data into the formats expected for:

1. Age (ndar\_subject01)
2. Sex at Birth (ndar\_subject01)
3. Patient Health Questionnaire-9 (PHQ-9, cde\_phq901 NDA data dictionary).

In addition, information from the Beck Depression Inventory will be deposited for all 155 research participants using the NDA bdi01 data dictionary. Deposited images will use the NDA image03 data dictionary. Data derived from the MRI images will be deposited into NDA using the imagingcollection01 data dictionary.

**Data Preservation, Access, and Associated Timelines**

Data will be deposited to NDA. Data will be findable through the NDA collection that will be established when this application is funded. The research community will have access to data at the end of the grant award. As required by NDA, studies that contain the data used for every publication will be created and shared when the manuscript is available as a pre-print. NDA studies have digital object identifiers (DOIs) to aid in findability. We will include the data DOI in relevant publications.

Researchers will request data using the standard processes at NDA, and the NDA data access committee will decide which requests to grant. The standard NDA data access process allows access for one year and is renewable. Once the data are submitted to NDA, that archive will control the long-term persistence of the data set.

**Access, Distribution, or Reuse Considerations**

The two existing data sets from NDA used consents that allow broad data sharing. The new dataset to be uploaded to NDA also was collected using informed consent terms that allow broad data sharing. Access to data housed by the NDA requires the completion of a [Data Use Certification](https://nda.nih.gov/faq.html#dac.3), which prohibits any redistribution or attempts to re-identify research participants.

For the 155 participants from our previous study, exact dates have been obscured via the Shift and Truncate method [1], which preserves within-case temporal relations.

**Oversight of Data Management and Sharing**

The Office of Sponsored Programs at University X has created a data management and sharing plan compliance system as part of their process for submitting the annual NIH progress report. That Office is collecting information related to the number of research participants that are deposited each reporting year. For this award, all of the data will be uploaded in the first year, so the data deposition oversight will end then. The Office of Sponsored Programs will look for the NDA data DOIs when papers are published and will include that information in the annual progress report.

**Validation Schedule**

Since this is a secondary data analysis application, validation of newly collected data will not occur. The new data to be deposited to NDA will go through their validation tool when the data are initially uploaded.

**Reference**

1. Hripcsak, G., Mirhaji, P., Low, A. F., & Malin, B. A. (2016). Preserving temporal relations in clinical data while maintaining privacy. *Journal of the American Medical Informatics Association*, *23*(6), 1040-1045.